REMARKS

Reconsideration of the present application in view of the above amendments and the following remarks is respectfully requested.

Claims 1, 2, 6-9 and 14-34 are pending, of which claims 1, 2, 14-27, 33 and 34 are under consideration.

Claim 1 has been amended to clarify the claimed subject matter without acquiescing to the rejections in the Office Action. Support for the amendments to claim 1 may be found, for example, at page 10, first paragraph, of the present application. No new matter has been added via the amendments to the claim.

The specification has been amended to add a sequence identification number for the sequence of the A motif of family A DNA polymerases or their Klenow fragments, DYSQIELR. A new sequence listing is provided herewith to add the missing sequence identification number for the above-noted sequence. Additionally, the application has been amended to insert the paragraph that incorporates the text copy of the Sequence Listing, as filed electronically via EFS-Web. Applicants respectfully submit that no new matter has been added via the amendments to the specification or to the sequence listing submitted herein. Applicants further submit that the above-identified application is now in compliance with 37 C.F.R. §§ 1.821-1.825 and WIPO Standard 25.

Election/Restrictions

Applicants respectfully request that because there is no prior art-based rejection against the elected species, the search and examination be extended to species other than the elected species, *i.e.*, a polymerase from *Thermus aquaticus*. In addition, as discussed below, Applicants believe that claim 1 and its dependent claims under consideration are allowable. Accordingly, Applicants further respectfully request that the withdrawn claims, which all ultimately refer to claim 1 and thus incorporate the features recited in claim 1, be examined in the present application.

Objection to Specification

The specification stands objected to because the amino acid sequence "DYSQIELR" recited in paragraph 6 on page 14 does not have a sequence identifier.

Applicants thank the Examiner for noting such an informality. Applicants have amended the specification to eliminate this informality and submitted a corrected sequence listing to include the sequence identifier for the above-noted sequence. Accordingly, Applicants submit that this ground of objection has been overcome. Withdrawal of this objection is respectfully requested.

Rejection Under 35 U.S.C. 112 (Written Description)

Claims 1, 2, 14-27, 33 and 34 stand rejected under 35 U.S.C. 112, first paragraph, as failing to meet the written description requirement. More specifically, the Office Action refers to the rejection in the previous Office Action, which is reproduced below:

> The specification, however, only provides the representative species of Taq and E. coli DNA polymerase mutants encompassed by these claims. There is no disclosure of any particular structure to function/activity relationship in the single disclosed species. The specification also fails to describe additional representative species of these family A DNA polymerases by any identifying structural characteristics or properties other than the activities in claims 1, for which no predictability of structure is apparent . . . . While applicants have functionally defined the claims in terms of the result of a motif C modification of a family A DNA polymerase, the claims are not limited structurally in any way, such that since the claims are drawn to a family A DNA polymerase which has been modified, the claims no longer have any structural limitations. Given this lack of species representative of such an unlimited genus of modified DNA polymerases, as encompassed by the claims, Applicants have failed to sufficiently describe the claimed invention, in such full, clear concise, and exact terms that a skilled artisan would recognize Applicants were in possession of the claimed invention.

Applicants respectfully traverse this ground of rejection. The written description requirement is met if the applicant conveys to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention. *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 1563-64; 19 USPQ 1111 (Fed. Cir. 1991). "A claim will not be invalidated on section 112 grounds simply because the embodiments of the specification do not contain examples explicitly covering the full scope of the claim language. This is because the patent specification is written

for a person of skill in the art, and such a person comes to the patent with the knowledge of what has come before. Placed in that context, it is unnecessary to spell out every detail of the invention in the specification; only enough must be included to convince a person of skill in the art that the inventor possessed the invention and to enable such a person to make and use the invention without undue experimentation." *Lizard-Tech, Inc. v. Earth Resource Mapping, PTY, Inc.*, 424 F.3d 1336, 1345; 76 USPQ2d 1724 (Fed. Cir. 2005).

In the response to the previous Office Action, Applicants argue that claim 1 of the present application is directed to a family A DNA polymerase or its Klenow fragment that has a modified motif C sequence and an enhanced mismatch discrimination as compared to the corresponding wild type polymerase or its Klenow fragment. In the modified motif C sequence, at least the amino acid residue Q879 in the wild type motif C sequence QVH in positions 879-881 of the *E. coli* DNA polymerase Klenow fragment shown in SEQ ID NO:2 has been replaced by a lipophilic amino acid residue. The term "family A DNA polymerase" is defined in the present application as referring to DNA polymerizing enzymes that contain the A motif with the sequence DYSQIELR in their active sites. Thus, the family A DNA polymerase or its Klenow fragment according to claim 1 of the present application has the following two structural features: (1) it has the A motif with the sequence DYSQIELR in its active site, and (2) it has a modified motif C sequence in which at least the amino acid residue Q879 in the wild type motif C sequence QVH in positions 879-881 of the *E. coli* DNA polymerase Klenow fragment has been replaced by a lipophilic amino acid residue. In addition, the family A DNA polymerase or its Klenow fragment claimed in the present application has the following two functional features: (1) it has DNA polymerase activity, and (2) it has an enhanced mismatch discrimination as compared to the corresponding wild type polymerase or its Klenow fragment.

In the previous response, Applicants further argued that one of ordinary skill in the art would not doubt that the present inventors had possession of the claimed invention at the filing of the present application. First, such a person would not doubt that the inventors had possession of a family A DNA polymerase. Applicants disagree with the assertion in the Office Action that the claims are not limited structurally in any way. As indicated above, because the present application defines the term "family A DNA polymerase" as referring to a DNA polymerase that contains the A motif with the sequence DYSQIELR in its active site, the claimed

family A DNA polymerase or its Klenow fragment has motif A with the sequence DYSQIELR in its active site. Applicants submit that family A DNA polymerases and its Klenow fragments were well known and characterized in the art at the filing of the present application. More specifically, as indicated in Patel *et al.* (J. Mol. Biol. 308:823-837, 2001, "Patel," copy enclosed), over 50 family A polymerases from different prokaryotic species had been sequenced (*see*, Abstract). Sequence alignments of these and other related polymerases show six conserved regions that form the active sites with regions 3, 4 and 5 corresponding to most conserved motifs A, B and C, respectively (*see*, pages 825 and 826, Figures 2 and 3). In addition, high-resolution crystal structures of several DNA polymerases had also been determined (*see*, page 824 and Figures 1 and 4-7). These studies in combination with mutagenesis analyses illustrated the structure-function relationship of family A DNA polymerases, including the function of individual amino acid residues in polymerization (*see*, pages 832-834). Thus, although not explicitly described in the present application, numerous family A DNA polymerases and their structure-function relationship were known in the art. In view of the advance understanding of family A DNA polymerases in the art at the time the present application was filed, one of ordinary skill in the art would not doubt that the present inventors had possession of a family A DNA polymerase other than Taq and *E. coli* DNA polymerases explicitly described in the present application.

Second, one of ordinary skill in the art would not doubt that the present inventors had possession of a family A DNA polymerase with an enhanced mismatch discrimination as compared to the corresponding wild type polymerase. As discussed above, the claimed DNA polymerase or its Klenow fragment has an additional structural feature: It has a modified motif C sequence in which at least the amino acid residue Q879 in the wild type motif C sequence QVH in positions 879-881 of the *E. coli* DNA polymerase Klenow fragment has been replaced by a lipophilic amino acid residue. The present application provides numerous exemplary substitutions of the QVH sequence (*see*, the third full paragraph on page 11). It further shows that certain exemplary substitutions of the QVH sequence resulted in enhanced mismatch discrimination activity of the modified DNA polymerases (*see*, Examples 3, 5 and 6, Figures 1-4). In view of such disclosure provided by the present application, Applicants submit that one of ordinary skill in the art would not doubt the possession of a family A DNA polymerase with an

enhanced mismatch discrimination as compared to the corresponding wild type polymerase by the present inventors.

In response to Applicants' previous arguments, it is asserted in the Office Action:

> It continues that applicants [sic] claims are drawn to any family A DNA polymerase or its Klenow fragment having a modified motif C sequence and an enhanced mismatch discrimination as compared to the corresponding wild type polymerase to is Klenow fragment, wherein in the modified motif C sequence at least the amino acid residue Q879 has been replaced with a lipophlic amino acid residue. Thus applicant's [sic] claims are clearly drawn to "mutant" family A DNA polymerases. Such is supported by applicant's specification at page 10, line 4, in which applicants state ". . . they also include thermostable DNA polymerases and their mutants."
>
> Thus while applicants submit that the term "family A DNA polymerase" is defined in the present application as referring to DNA polymerizing enzymes that contain the A motif with the sequence DYSQIELR in their active site, applicants claims are drawn to "mutant family A DNA polymerases" for which the structural limitations associated with applicants referred to "family A DNA polymerases" are not necessarily associated with the claimed mutant family A DNA polymerases". Thus it remains that applicants [sic] only structural limitations refer to the mutated amino acid position and not to the remainder of the "mutated family A DNA polymerase".
>
> Thus the functional features of having DNA polymerase activity, and especially an enhanced mismatch discrimination activity are not associated with any structural features including the motif A with the sequence DYSQIELR in its active site. While the art may describe family A DNA polymerases (Patel *et al.* (J. Mol. Bio. 308:823-837, 2001, "Patel," copy enclosed), applicants claims are not so limited.

Applicants disagree with the interpretation of the claims in the Office Action. Applicants submit that one of ordinary skill in the art, in view of the present application (especially the explicit definitions of "family A DNA polymerases" and "DNA polymerase according to the invention" provided in the application), would interpret that the claimed family A DNA polymerase or its Klenow fragment has the A motif with the sequence DYSQIELR and DNA polymerase activity. More specifically, the present application provides:

> "Family A DNA polymerases" (also referred to as "polymerases I") are those DNA-polymerizing enzymes which contain the A motif with the sequence DYSQIELR in their active site. They also include the enzymes described herein which have mutations in the C motif. In particular, they also include thermostable DNA polymerases and their mutants. (*see*, first full paragraph on page 10)

The present application further provides:

> A "DNA polymerase according to the invention" is a family A DNA polymerase as defined above which includes the A motif with the sequence DYSQIELR in its active site and comprises particular mutations in the C motif. In particular, they also include thermostable DNA polymerases with mutations in the C motif. These mutations are conservative substitutions of the QVH amino acid residues of the C motif and/or the above defined non-conservative substitutions. (*see*, sixth paragraph on page 14)

The interpretation in the Office Action that the Family A DNA polymerase claimed in the present application may contain mutations in the A motif directly contradicts the above explicit definitions, and is thus incorrect. Although Family A DNA polymerases may include thermostable DNA polymerase mutants, one of ordinary skill in the art would know, based on the definition provided in the present application, that such mutants do not have mutations in the A motif, but in the C motif.

Nevertheless, to facilitate allowance and to avoid misinterpretation of the claims, Applicants have amended claim 1 to specifically recite that the family A DNA polymerase or its Klenow fragment comprises the A motif with the sequence DYSQIELR. Accordingly, as indicated in Applicants' previous response, the family A DNA polymerase or its Klenow fragment according to claim 1 of the present application has the following two structural features: (1) it has the A motif with the sequence DYSQIELR in its active site, and (2) it has a modified motif C sequence in which at least the amino acid residue Q879 in the wild type motif C sequence QVH in positions 879-881 of the *E. coli* DNA polymerase Klenow fragment has been replaced by a lipophilic amino acid residue. In addition, the family A DNA polymerase or its Klenow fragment claimed in the present application has the following two functional features: (1) it has DNA polymerase activity, and (2) it has an enhanced mismatch discrimination as compared to the corresponding wild type polymerase or its Klenow fragment. Moreover, also as indicated in Applicants' previous response, family A DNA Polymerases and their Klenow fragments were well known and characterized in the art at the filing of the present application (*see*, Patel, previously submitted). Thus, one of ordinary skill in the art would not doubt that the present inventors had possession of family A DNA polymerases other than Taq and *E. coli* DNA polymerases explicitly described in the present application. Furthermore, the present application

provides numerous exemplary substitutions of the QVH sequence in the C motif (see, the third full paragraph on page 11) and experimental data showing that certain exemplary substitutions resulted in an enhanced mismatch discrimination activity of the modified DNA polymerases (*see*, Examples 3, 5 and 6, Figures 1-4). Accordingly, one of ordinary skill in the art would not doubt the possession of the family A DNA polymerase with an enhanced mismatch discrimination activity claimed in the present application by the present inventors at the time of the present application.

In view of the above remarks, Applicants submit that this ground of rejection under 35 U.S.C. 112, first paragraph, has been overcome. Applicants respectfully request that this rejection be withdrawn.

Rejection Under 35 U.S.C. 112 (Enablement)

Claims 1, 2, 14-27, 33 and 34 stand rejected under 35 U.S.C. 112, first paragraph, as not enabled. More specifically, it is asserted in the Office Action that the specification, while enabling for a family A DNA polymerase comprising the amino acid sequence of SEQ ID NO:2 in which the amino acid residue at position Q879 has been replaced with a lipophilic amino acid residue, does not reasonably provide enablement for any family A DNA polymerases that have a modified motif C sequence and an enhanced mismatch discrimination as compared to a corresponding wild type polymerase or its Klenow fragment in which at least the amino acid residue Q879 in the motif C sequence QVH at positions 879-881 based on the *E. coli* DNA polymerase Klenow fragment has been replaced by a lipophilic amino acid residue.

Applicants respectfully traverse this ground of rejection. It seems that the non-enablement rejection in the Office Action is primarily based on the assertion that the present claims place minimal if any structural limits on the claimed modified polymerases. As discussed above, Applicants respectfully disagree with this assertion. However, to facilitate allowance and without acquiescing to the rejection in the Office Action, Applicants have amended claim 1 to specifically recite that the family A DNA polymerase or its Klenow fragment comprises the A motif with the sequence DYSQIELR. Thus, the claimed DNA polymerase has the following two structural features: (1) it has the A motif with the sequence DYSQIELR in its active site, and (2) it has a modified motif C sequence in which at least the amino acid residue Q879 in the wild type

motif C sequence QVH in positions 879-881 of the *E. coli* DNA polymerase Klenow fragment has been replaced by a lipophilic amino acid residue. The first structural feature relates to the DNA polymerase activity of the claimed modified family A DNA polymerase, and the second structural feature relates to the enhanced mismatch discrimination activity of the claimed modified family A DNA polymerase. As discussed in detail below, one of ordinary skill in the art would know how to make (and use) the claimed DNA polymerase in view of the present application in combination with knowledge in the art available when the present application was filed.

In the previous response, Applicants argued that one of ordinary skill in the art would know how to make (and use) a family A DNA polymerase based on the knowledge available at the filing of the present application. As discussed above, family A DNA polymerases were well known, and their structure-function relationship was well characterized. For example, as described in Patel, family A DNA polymerases have 6 conservative regions that form their active sites, and the tertiary structures of several family A DNA polymerases were characterized. In addition, function of individual amino acids was further studied, which shows that very few (<10) amino acid residues within the highly conserved motifs A, B and C have a direct role during nucleotide binding and incorporation, and only those residues that are important during catalysis and/or for protein folding need to be maintained, while all other residues are mutable (*see*, the Conclusions section on page 835, the first full paragraph in the right column on page 833 and the first paragraph in the right column on page 834, and the second paragraph in the left column on page 835). Thus, both the regions of family A DNA polymerases that may be modified without affecting polymerase activities and the great tolerance of this type of polymerase to modification were known in the art. In view of such knowledge in the art, one of ordinary skill in the art would be able to make family A DNA polymerases, including both naturally occurring family A DNA polymerases and their variants with retained DNA polymerase activity.

In response, it is asserted that

> While one of ordinary skill in the art would know how to make (and use) a family A DNA polymerase based on the knowledge available at the filing of the present application, as discussed above, the breadth of the claimed mutant family A DNA polymerases is considerably broader. As discussed above those structural

> limitations associated with a family a [sic] DNA polymerase are not necessarily associated with a mutant family A DNA polymerase.
>
> It is this breadth of the claimed mutant DNA polymerases and the lack of guidance associated with "enhanced mismatch discrimination activity" associated with such a broad genus of mutant DNA polymerases that results in the lack of scope of enablement. While applicants argue that those of skill in the art would know of certain associated structures or domains associated with the claimed DNA polymerases, such are not necessarily associated with the claimed mutant DNA polymerases.

As indicated above, to facilitate allowance and without acquiescing to the rejection in the Office Action, Applicants have amended claim 1 to specify that the family A DNA polymerase or its Klenow fragment claimed in the present application comprise the A motif with the sequence DYSQIELR. Thus, the above-noted structural limitations associated with a family A DNA polymerase are associated with the family A DNA polymerase or its Klenow fragment claimed in the present application.

In the previous response, Applicants further argued that the reference cited in the Office Action, Ngo, is insufficient to support the non-enablement rejection. First, this reference was published in 1994, 10 years earlier than the priority date of the present application. Thus, it does not describe the state of the art at the time of the present application. Second, Ngo states that it is not known whether there exists an efficient algorithm for predicting the structure of a given protein from its amino acid sequence alone (*see*, second full paragraph on page 492). However, computer prediction via algorithms is not the only way that the structure of a given protein may be analyzed. As indicated in Patel, high-resolution crystal structures of several polykaryotic DNA polymerases have been determined, and the structure and function relationship of these enzymes have been characterized by various means, including sequence alignments, crystal structural analysis, and mutagenesis studies. Thus, Ngo does not specifically describe the state of the art regarding polykaryotic DNA polymerases.

In response to Applicants' previous arguments, it is asserted in the Office Action:

> While applicants argue that the reference Ngo et al. is not representative of the skill in the art at the time of filing, and that much has been learned since the publishing of the reference Ngo et al., applicants [sic] reference to the teaching of Patel et al. while helpful is not sufficient given the extreme breadth of the claimed genus of "mutant" DNA polymerases.

Again, as indicated above, Applicants have amended the claim to specify that the family A DNA polymerase or its Klenow fragment claimed in the present application comprise the A motif with the sequence DYSQIELR. Thus, the scope of the claims in the present application is not as broad as asserted in the Office Action. Applicants submit that the claims as currently pending are enabled by the present application in view of the advanced state of the art at the time of filing of the present application.

In the previous response, Applicants also argued that one of ordinary skill in the art would also know how to modify a family A DNA polymerase to increase its mismatch discrimination activity in view of the present application. The present application provides that mismatch discrimination activity of a family A DNA polymerase may be enhanced by replacing Q879 in the motif C sequence QVH with a lipophilic amino acid residue, such as Gly, Ala, Val, Leu and Ile (*see*, *e.g.*, second paragraph on page 5 and third full paragraph on page 11). Example 1 of the present application teaches how to make and purify Klenow fragment variants with substitution in the motif C sequence QVH. Example 2 of the present application teaches how to screen the Klenow fragment variants for those with an enhanced mismatch discrimination activity. Examples 3, 5 and 6 demonstrate that exemplary variants containing substitutions in the motif C sequence QVH have enhanced mismatch discrimination activity in primer extension assays and real time PCR experiments. In view of the teachings of the present application, one skilled in the art would be able to make a family A DNA polymerase variant having Q879 replaced with a lipophilic amino acid residue (*see*, Example 1) and determine its mismatch discrimination activity (*see*, Examples 2, 3, 5 and 6) without undue experimentation.

In response to Applicants' previous arguments, it is asserted in the Office Action:

> Finally while one of skill in the art would know mechanistically how to modify a family A DNA polymerase to increase its mismatch discrimination activity in view of the present application, it remains that the art and application do not provide guidance at to the specific of exactly how to modify a family A polymerase to increase its mismatch discrimination activity.

Applicants disagree with the above assertion. As indicated above, the present application provides that mismatch discrimination activity of a family A DNA polymerase may be enhanced by replacing Q879 in the motif C sequence QVH with a lipophilic amino acid residue, such as Gly, Ala, Val, Leu and Ile and demonstrates enhanced mismatch discrimination

activity of exemplary family A DNA polymerases. Applicants submit that contrary to the assertion in the Office Action, the present application provides specific guidance for one of ordinary skill in the art to make and use the claimed family A DNA polymerase without undue experimentation and thus enables the claims currently pending in the present application.

In view of the above remarks, Applicants submit that this ground of rejection under 35 U.S.C. 112, first paragraph, has been overcome. Withdrawal of this rejection is respectfully requested.

Applicants believe that the claims pending in the present application are now allowable. Favorite consideration and a Notice of Allowance are earnestly solicited.

The Director is authorized to charge any additional fees due by way of this Amendment, or credit any overpayment, to our Deposit Account No. 19-1090.

Respectfully submitted,

SEED Intellectual Property Law Group PLLC

/Qing Lin/
Qing Lin, Ph.D.
Registration No. 53,937

QXL:kw

Enclosure: Patel *et al.* (J. Mol. Biol. 308:823-837, 2001)

701 Fifth Avenue, Suite 5400
Seattle, Washington 98104
Phone: (206) 622-4900
Fax: (206) 682-6031

1808346_1.DOC

doi:10.1006/jmbi.2001.4619 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 308, 823–837

JMB

REVIEW ARTICLE

Prokaryotic DNA Polymerase I: Evolution, Structure, and "Base Flipping" Mechanism for Nucleotide Selection

Premal H. Patel[1], Motoshi Suzuki[1], Elinor Adman[1], Akeo Shinkai[1] and Lawrence A. Loeb[1]*

[1]*The Joseph Gottstein Memorial Cancer Laboratory, Department of Pathology, University of Washington School of Medicine Seattle, WA 98195-7705, USA*

**Corresponding author*

Accurate transmission of DNA material from one generation to the next is crucial for prolonged cell survival. Following the discovery of DNA polymerse I in *Escherichia coli*, the DNA polymerase I class of enzymes has served as the prototype for studies on structural and biochemical mechanisms of DNA replication. Recently, a series of genomic, mutagenesis and structural investigations have provided key insights into how Pol I class of enzymes function and evolve. X-ray crystal structures of at least three Pol I class of enzymes have been solved in the presence of DNA and dNTP, thus allowing a detailed description of a productive replication complex. Rapid-quench stop-flow studies have helped define individual steps during nucleotide incorporation and conformational changes that are rate limiting during catalysis. Studies in our laboratory have generated large libraries of active mutant enzymes (8000) containing a variety of substitutions within the active site, some of which exhibit altered biochemical properties. Extensive genomic information of Pol I has recently become available, as over 50 *polA* genes from different prokaryotic species have been sequenced. In light of these advancements, we review here the structure-function relationships of Pol I, and we highlight those interactions that are responsible for the high fidelity of DNA synthesis. We present a mechanism for "flipping" of the complementary template base to enhance interactions with the incoming nucleotide substrate during DNA synthesis.

Keywords: DNA polymerase; structure-function; replication; active site; recombination

Introduction

Primary functions and sequence of prokaryotic DNA Pol I

Prokaryotic species exist in diverse climates, ranging from radiation-resistant *Deinococcus radiodurans* (believed to have evolved one billion years ago; Battista, 1997), to thermostable *Thermus aquaticus* (first isolated from hot springs in Yellow Stone National Park; Chien *et al.*, 1976) to enteric bacteria *Escherichia coli*, to obligate intracellular bacteria *Chlamydia trachomatis*. Despite this environmental diversity, prokaryotic DNA Pol I sequence has remained remarkably conserved through at least one billion years (Gutman *et al.*, 1993; Joyce *et al.*, 1982). The gene encoding DNA polymerase I (*polA*) contains approximately 3000 base-pairs that encodes approximately 1000 amino acid residues in a simple polypeptide chain. Organisms separated by a billion years of evolution (e.g. *Deinococcus-Thermus* genera and *E. coli*) have DNA Pol I groups that exhibit similar activities, retain ≈35% amino acid identity and ≈50% homology. Interestingly, the lethal effects of deleting endogenous Pol I activity from a specific prokaryote can be complemented by Pol I from another organism. For example, *Taq* Pol I can fully restore viability of *E. coli* lacking endogenous Pol I activity (Suzuki

Abbreviations used: Pol I, polymerase I; RT, reverse transcriptase.

E-mail address of the corresponding author: laloeb@u.washington.edu

0022-2836/01/050823–15 $35.00/0

et al., 1996a; Patel & Loeb, 2000a), and *E. coli* Pol I can restore viability in *D. radiodurans* lacking endogenous Pol I (Gutman *et al.*, 1994). This interchangeability suggests that Pol I functions in similar pathways in diverse prokaryotes.

Biochemical and genetic investigations in *E. coli* have contributed to our understanding of Pol I's role *in vivo*. *E. coli* Pol I, the first polymerase discovered, is the most abundant polymerase within its host (approximately 400 molecules per cell) and functions primarily to fill DNA gaps that arise during DNA repair, recombination and replication (Kornberg & Baker, 1992). Prokaryotic Pol Is have two functional domains located on the same polypeptide: a 5′-3′ polymerase (located towards the C terminus) and 5′-3′ exonuclease (located towards the N terminus) (Joyce & Grindley, 1984). *E. coli* Pol I contains in addition a "non-essential" proofreading 3′-5′ exonuclease domain which functions to excise DNA replication errors. DNA Pol Is from diverse species within the genera *Rickettsia* and *Thermus* do not contain motifs ExoI, ExoII, and ExoIII, which are necessary for 3′-5′ exonuclease function. Pol I uses its 5′ nuclease activity to remove the ribonucleotide portion of newly synthesized Okazaki fragments and DNA polymerase activity to fill in the resulting gap. *polA(−) E. coli* join Okazaki fragments tenfold slower than wild-type bacteria, and are sensitive to DNA damage by UV damage and alkylating agents (Deluca & Cairns, 1969). During repair, Pol I fills in DNA gaps that result from the removal of a variety of DNA lesions (e.g. UV-induced thymidine dimer, the oxidative lesion 8-oxo guanine, and the alkylation lesion 4-methyl adenine, etc.; for a review of repair mechanisms, see Friedberg *et al.*, 1995).

The Pol I family has served as a prototype for studying DNA polymerase mechanisms. There are at least six known families of DNA polymerases, which are grouped based on amino acid sequence homologies: A, B, X, RT, Pol III, and UmuC/DinB families. Family A polymerases are found primarily in organisms related to prokaryotes and include prokaryotic DNA polymerse I, mitochondrial polymerase γ, and several bacteriophage polymerases including those from odd-numbered phage (T3, T5, and T7). Family B polymerases are present in bacteriophages, viruses, archea and eukarotes. Many of these polymerses function to replicate the host genome, and include those from even-numbered phages T4 and T6, herpes viruses, archeal pol "Vent", and mammalian pols α, δ and ε. Family X, containing mammalian pol β, λ, and μ, function during DNA repair. The reverse transcriptase family contains RTs from retroviruses as well as eukaryotic telomerases (Lingner *et al.*, 1997). The prokaryotic Pol III family encompasses the DNA polymerases that replicate the majority of bacterial genomes. The recently discovered UmuC/DinB family includes pols η, i, κ, and deoxycytidyl transferase (for a review, see Friedberg *et al.*, 2000).

Three-dimensional structures of the Pol I active site

High-resolution crystal structures of DNA polymerases in the Pol I family including *Taq* Pol I (Kim *et al.*, 1995; Li *et al.*, 1998), *E. coli* Pol I (Klenow) (Beese *et al.*, 1993a; Ollis *et al.*, 1985), *Bacillus* polymerase (Klenow-like fragment) (Kiefer *et al.*, 1998) and bacteriophage T7 DNA polymerase (Doublie *et al.*, 1998) have been determined (for a review, see Steitz, 1999). The X-ray crystal structures of these polymerases resemble in overall morphology a cupped human right hand, with fingers (which bind an incoming nucleotide and interact with the single-stranded template), palm (which harbors the catalytic amino acid residues and also binds an incoming dNTP) and thumb (which binds double-stranded DNA) subdomains. Distantly related prokaryotic polymerases within the Pol I family such as *Taq* Pol I and *E. coli* Pol I exhibit less extensive homologies, nevertheless have three-dimensional structures that are virtually identical, with secondary structural elements, α-helices and β-strands, that are all nearly completely superimposible (Figure 1). These studies show extensive sequence heterogeneity between enzymes can produce identical folding patterns. The structural identity in turn suggests Pol I enzymes from

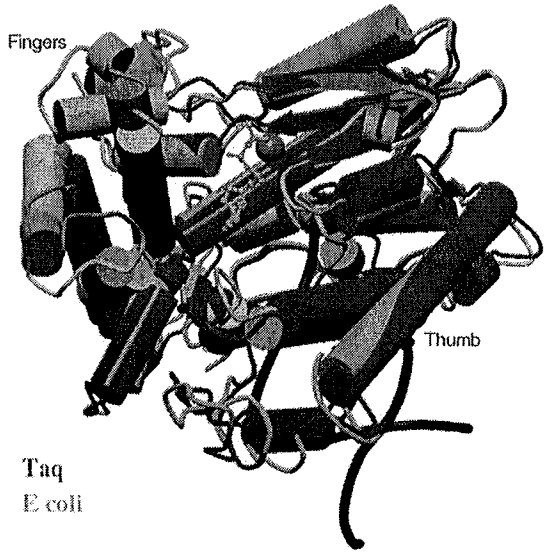


Figure 1. Superposition of *Taq* Pol I (Li *et al.*, 1998) and *E. coli* Pol I (Klenow fragment; Beese *et al.*, 1993a) crystal structures. *Taq* Pol I (red) and *E. coli* Pol I (blue) are separated in evolution by one billion years. The structures of these two enzymes are superimposible, such that corresponding α-helices and β-strands adopt nearly identical folding patterns (on average, <1 Å difference). Conformation of the fingers subdomain (magenta and teal) differ slightly, suggesting this region is flexible. Superpostioning of other related Pol I class of enzymes yielded similar results. All structure illustrations were prepared by E. Adman using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt & Murphy, 1994).

diverse organisms function to incorporate nucleotides by identical mechanisms.

Amino acid sequence alignments of related enzymes can help identify important regions with common catalytic functions (Figure 2). Delarue *et al.* (1990) compared the C-terminal polymerase domains amongst distantly related members of the Pol I family of enzymes. Following alignment, five conserved regions (numbered 1-5) within diverse DNA Pol Is were delineated; in addition, we identify a sixth conserved region (region 6). Three of these regions (3, 4, and 5) resembled the most highly conserved domains of mammalian DNA pol α in family B (termed motif A, B, and C; Figures 3 and 4). Analysis of high-resolution crystal structures of family A polymerases in complex with DNA and an incoming nucleotide suggests each of these six regions have an important role during DNA synthesis (Figure 2, region 5). Region 1 amino acid residues are located at the tip of the thumb subdomain and form a helix-loop that interacts with the minor groove of the double-stranded DNA during the nucleic acid binding step. Region 2 amino acid residues are located within the palm subdomain and interact with the template strand (along the minor groove) and thus, form the "template grip". Region 6 amino acid residues interact with the first template base; as a result, this template could be "flipped out" of the helix axis by $\geqslant 90°$, such that it cannot base-pair with the incoming dNTP (see below). Amino acid residues within regions 3, 4, and 5 correspond to motifs A, B, and C, respectively, and are located near the incoming nucleotide triphosphate and thus form

```
Taq  : 444 MEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVL---------FDELG
Bst  : 488 MEFTGVKVDTKRLEQMGAELTEQLQAVERRIYELAGQEFNINSPKQLGTVL---------FDKLQ
                                                FN +S +++    L          + +
T7   :     ****************************************FNPSSRDHIQKKLQEAGWVPTKYTDKG
                                                    ————1———

Taq  : 500 LPAIGKTEKTGKR---STSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLH
Bst  : 544 LPVLKKT-KTGY----STSADVLEKLAPHHEIVEHILHYRQLGKLQSTYIEGLLKVVHPVTGKVH
           +P++ +    G          +++ +     I + I +  +  K+   Y+           G++H
T7   : 361 APVVDDEVLEGVRVDDPEKQAAIDLIKEYLMIQKRIGQSAEGDKAWLRYV--------AEDGKIH

                                                                ————————————
Taq  : 562 TRFNQTATATGRLSSSDPNLQNIP-VRTPLGQRIRRAFIAE--------EGWLLVALDYSQIELR
Bst  : 604 TMFNQALTQTGRLSSVEPNLQNIP-IRLEEGRKIRQAFVPSE-------PDWLIFAADYSQIELR
             N     TGR++   PNL  IP +R   G++ R AF              W+ + +D+S +ELR
T7   : 418 GSVNPNGAVTGRATHAFPNLAQIPGVRSPYGEQCRAAFGAEHHLDGITGKPWVQAGIDASGLELR
             ————————2————————                                ————————A————

                                                    ————————————
Taq  : 618 VLAHLSGDENLIRVFQE--GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQ
Bst  : 661 VLAHIAEDDNLIEAFRR--GLDIHTKTAMDIFHVSEEDVTANMRRQAKAVNFGIVYGISDYGLAQ
           LAH   + +  +   +    DIHT+         +  +   +R+ AK   +G LYG     + Q
T7   : 483 CLAHFMARFDNGEYAHEILNGDIHTKN--------QIAAELPTRDNAKTFIYGFLYGAGDEKIGQ
                                                    ————————B————

Taq  : 681 ELAIPYEEAQAFIERYFQSFPKV---RAWIEKTL-EEGRRRGYVETLFGRRRYVPDLEARVKSVR
Bst  : 724 NLNITRKEAAEFIERYFASFPGV---KQYMDNIV-QEAKQKGYVTTLLHRRRYLPDITSRNFNVR
           + +  ++     +++    P +   +   ++ + +  +       +  +RR++  +  R   VR
T7   : 540 IVGAGKERGKELKKKFLENTPAIAALRESIQQTLVESSQWVAGEQQVKWKRRWIKGLDGRKVHVR

                                                            ————————————
Taq  : 742 EAAERMAFNMPVQGTAADLMKLAMVKLF-PRLEEMG-AR--------MLLQVHDELVLEA-PKER
Bst  : 785 SFAERTAMNTPIQGSAADIIKKAMIDLSVSVREERLQAR--------LLLQVHDELILEA-PKEE
             +   A N+ +Q   A + K   ++       E   ++          + VHDE+ +    +E
T7   : 605 --SPHAALNTLLQSAGALICKLWIIKTE-EMLVEKG-LKHGWDGDFAYMAWVHDEIQVGCRTEEI
                 ————————6————————                            ————————C————

Taq  : 796 AEAVARLAKEVMEGV---YPLAVPLEVEVGIGEDWLSAK* 831
Bst  : 841 IERLCRLVPEVMEQA---VALRVPLKVDYHYGPTWYDAK* 876
           ++ + + + E+M+ +         L+ +   G  W
T7   : 666 AQVVIETAQEAMRWVGDHWNFRCLLDTEGKMGPNWAICH* 700
```

Figure 2. DNA polymerase I sequence alignments. Sequences of *Taq* Pol I, T7 Pol and *Bst* Pol I (accession numbers 70368, GI2126855) were obtained from GenBank and aligned using NCBI **BLAST** family of programs. Conserved amino acid residues are listed, and conserved motifs are highlighted. Regions of insertion or deletions are marked by (–), and (*) indicates an inserted segment within T7 DNA polymerase that allows binding with an accessory protein and processivity factor, thioridoxin.

	MOTIF A	MOTIF B	MOTIF C
THERMUS AQUATICUS	LLVALDYSQIELR	RRAAKTINFGVLY	LLQVHDELVLE
THERMUS THERMOPHILUS	ALVALDYSQIELR	RRAAKTVNFGVLY	LLQVHDELLLE
THERMUS FILIFORMIS	LLLAADYSQIELR	RRAAKTVNFGVLY	LLQVHDELVLE
DEINOCOCCUS RADIODURANS	TLIAADYSQIELR	RRAAKTVNFGVLY	LLQVHDELLIE
ESCHERICHIA COLI	VIVSADYSQIELR	RRSAKAINFGLIY	IMQVHDELVFE
HAEMOPHILUS INFLUENZAE	SIVAADYSQIELR	RRNAKAINFGLIY	IMQVHDELVFE
STREPTOCOCUS PNEUMONIAE	VLLSSDYSQIELR	RRNAKAVNFGVVY	LLQVHDEIVLE
MYCOBACTERIUM TUBERCULOSIS	ELMTADYSQIEMR	RRRVKAMSYGLAY	LLQVHDELLFE
MYCOBACTERIUM LAPRAE	ELMTADYSQIEMR	RRRVKAMSYGLAY	LLQVHDELLFE
TREPONEMA PALLIDUM	ELISADYTQIELV	RRIAKTINFGIVY	LLQVHDELIFE
CHLAMYDIA TRACHOMATIS	YFLAADYSQIELR	RYQAKAVNFGLVY	LLQIHDELLFE
BORIELA BURGDORFERI	IFISADYSQIELA	RRIAKSINFGIIY	LLQVHDEMLIE
HELICOBACTER PYROLI	CLLGVDYSQIELR	RSIAKSINFGLVY	LLQVHDELIFE
LACTOCOCCUS LACTIS	LLLSSDYSQIELR	RRNAKAVNFGVVY	LLQVHDEIILD
MYTHELOBACTERIUM	KLISADYSQIELR	RRRAKTINFGIIY	LLQVHDELVFE
RHODOTHERMUS OBAMENSIS	KLLSADYVQIELR	RRRAKMVNYGIPY	LLQVHDELVFE
RICKETTSIA PROWAZEKII	KLISADYSQIELR	RRKAKAINFGIIY	ILQIHDELLFE
STREPTOMYCES COELICOLOR	SLMTADYSQIELR	RRKIKAMSYGLAY	LLQVHDEIVLE
BACILLUS STEAROTHERMOPHILUS	LIFAADYSQIELR	RRQAKAVNFGIVY	LLQVHDELILE
SYNECHOCYSTIS SP	LLVSADYSQIELR	RNLGKTINFGVIY	LLQVHDELIFE
AQUIFEX AEOLICUS	TFVISDFSQIELR	RQLAKAINFGLIY	VNLVHDEIVVE
APSE-1 DNA polymerase	klvisdlsniegr	rqigkvmelglgy	ivtvhdeiise
T7 DNA polymerase	vqagidasglelr	rdnaktfiygfly	mawvhdeiqvg
T5 DNA polymerase	rviawdlttaevy	rqaakaitfgily	vmlvhdsvvai
Concensus sequence	xhhhhDhhxhEhx	RpxxKxxxhGhhY	hhxhHDxhxxx

Figure 3. Pol I sequences of motifs A, B, and C. Sequences of prokaryotic Pol I (in capital letters) and bacteriophage Pols (in lowercase letters) were obtained from GenBank and aligned using NCBI **BLAST** family of programs. The sequences of conserved motifs A, B, and C are shown. In the consensus sequence (bottom), amino acids that are absolutely conserved are listed; h represents hydrophobic amino acid, and x represents diverse amino acids occupy a particular position.

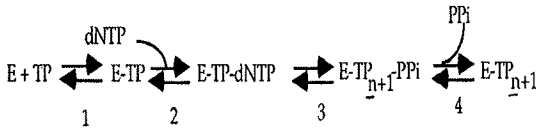
portions of the dNTP binding cleft within the DNA polymerase active site.

Motifs A and C are located within the palm subdomain and are conserved in structure within families A, B, RT, and X polymerases, while motif B is located in the fingers subdomain. Structurally, motifs A, B and C of *Taq* DNA Pol I, *E. coli* Pol I, *Bst* Pol I, and T7 DNA Pol are nearly indistinguishable (Figures 3 and 4). Motif A begins at an antiparallel β-strand containing predominantly hydrophobic residues and continues to an α-helix. Amino acid sequences of a large variety of naturally occurring prokaryotic DNA Pol Is (Figures 3 and 4) show many sequences can produce the "antiparallel β-strand" portion of motif A; however, the "turn and α-helix" portions consistently contain the "DYSQIELR" amino acid sequence. Motif B constitutes an α-helix located in the fingers subdomain and is found in both A and B families of DNA polymerases. DNA Pol I sequence alignment indicates motif B contains a consensus sequence RRxhKhhNFGhhY, where h represents hydrophobic amino acid and x indicates any amino acid. Structurally, motif C amino acids form two antiparallel β-strands. While diverse amino acid sequences form the antiparallel β-strands, the loop between these strands invariably contains the sequence "HDE". Comparison of DNA polymerase sequences from bacteriophage, which evolve at very high rates, with that of prokaryotic Pol Is shows that few active site amino acid residues are conserved between these groups. For example, only motif A amino acid residues Asp610 and Glu625 and motif B residues Arg659, Lys663, Gly668, and Tyr671, and motif C amino acid residues His784 and Asp785 are entirely conserved. (All amino acid numbering is based on the *Taq* Pol I sequence.) Sequence alignments between distant family members indicate only a few amino acid residues within the polymerase active site are conserved in the Pol I family of enzymes; however, a comparison of individual Pol I sequences to crystal structures indicates that a significant level of sequence heterogeneity (>50 %) can yield structural identity. To further understand the potential roles of active site amino acid residues and gain insights into DNA polymerization mechanisms, we will examine structures of Pol I complexed with diverse substrates encountered during DNA replication.

Pathway of DNA Synthesis

The rates and steps involved during single nucleotide incorporation reactions have been determined for many polymerases, including those in the Pol I class (e.g. *Taq* Pol I (Brandis *et al.*, 1996),

T7 DNA Pol (Patel *et al.*, 1991), and *E. coli* Pol I (Bryant *et al.*, 1983; Kuchta *et al.*, 1988)). There emerges a general mechanism for single nucleotide additions:



(1)

During DNA synthesis (equation (1)): 1, polymerase (E) binds with template-primer (TP); 2, the appropriate dNTP binds with polymerase-DNA complex; 3, a nucleophilic attack results in phosphodiester bond formation; and 4, pyrophosphate (PPi) is released. Rapid quench kinetic experiments indicate that DNA binding (1) and nucleotide binding (2) occur very rapidly. The rate-limiting step is either phosphodiester bond formation or a conformational change that precedes nucleotide incorporation. The incorporation of complementary nucleotides involves dynamic interactions between the polymerase with its nucleic acid and dNTP substrates. Polymerases are proposed to undergo at least four significant conformational changes: (i) during the DNA binding step; (ii) subsequent to the dNTP binding step and immediately preceding chemical catalysis; (iii) subsequent to nucleotide incorporation during PPi release; and during (iv) translocation towards the new primer 3′-OH terminus. Theoretical studies indicate that multiple conformational changes contribute to the fidelity of catalysis by DNA polymerases (Beckman & Loeb, 1993). High resolution crystal structures exist within the Pol I family of enzymes with all of these intermediates: (1) without DNA and dNTP (*Taq* Pol I, *E. coli* Pol I (Klenow)); (2) with DNA in the active site but without dNTP (*Taq* Pol I); (3) with DNA and dNTP (*Taq* Pol I and T7 pol); (4) with pyrophosphate (Klenow)(Beese *et al.*, 1993b); and (5) translocated following dNTP addition (*Bst* Pol I). Of these crystal structures, *Taq* Pol I:DNA:dNTP (Li *et al.*, 1998), T7:DNA:dNTP (Doublie *et al.*, 1998) and *Bst* Pol:DNA:dNTP (Kiefer *et al.*, 1998) are particularly revealing, as each of these high-resolution structures depicts important steps during catalysis.

Binding of polymerase to DNA

The first step in polymerization involves the association of the polymerase with the template-primer. Comparisons amongst the crystal structures of *Taq* Pol I, *Bst* Pol I, and *E. coli* Pol I in complex with DNA show that the thumb subdomain changes conformation to nearly completely wrap around the DNA. Two significant enzyme conformational changes occur within *Taq* Pol I thumb subdomain (Li *et al.*, 1998): (1) the thumb subdomain rotates towards the palm subdomain; and (2) the conserved amino acid residues located within the tip of the thumb domain (helices H1 and H2; region 1; Figures 2 and 4) rotate in the opposite direction relative to the rest of the thumb such that the tip is in proximity to the DNA. Together, these changes result in an approximately 30 Å wide cylinder that almost completely engulfs the DNA, such that conserved amino acid residues within the tip of the thumb subdomain (region 1) "grip" the DNA along the minor groove. A conformation change within the thumb subdomain following DNA binding has also been reported for HIV-1 RT (Patel *et al.*, 1995). In each of the diverse polymerase:DNA high-resolution complexes studied, the enzyme interacts primarily with the sugar-phosphate DNA backbone along the minor groove, and these interactions are associated with bending of the DNA such that it adopts an S-shaped conformation. *Taq* Pol I interactions with the thumb subdomain along the minor groove result in bending double-stranded DNA three bases from the primer terminus, causing the DNA near the active site to adopt an *A*-like conformation (Figures 4 and 5). A second set of interactions, where the single strand template interacts with the palm subdomain (by region 2 or "template grip" amino acid residues), also contributes to this bend. A third set of polymerase:DNA interactions, results in a $\geqslant 90°$ rotation of the first template base and a 180° rotation of the adjacent template bases, such that the template bases are flipped outside the DNA helix axis and away from the dNTP binding site (Figures 4 and 5). This unusual S-shaped DNA conformation, in which the first bend is induced by interactions with the thumb and palm subdomains and the second bend is induced by portions of the active site interacting with the single-strand template, is also seen in high-resolution structures of *Bst* Pol I, T7 pol, and HIV-1 RT bound to DNA (Figure 4). Thus, this DNA conformation, including template base initially flipped out of the helix axis by $\geqslant 90°$, may have a general relevance to the DNA polymerization mechanism by a wide variety of enzymes.

Nucleotide binding

A second conformation change within polymerases occurs during the dNTP-binding step. High-resolution crystal structures of *Taq* Pol I (Li *et al.*, 1998), T7 DNA Pol (Doublie *et al.*, 1998), HIV-1 RT (Huang *et al.*, 1998), and *Bst* Pol I (Kiefer *et al.*, 1998) bound to DNA in the presence of an incoming nucleotide suggest at least three events occur during the dNTP binding step (Figures 5 and

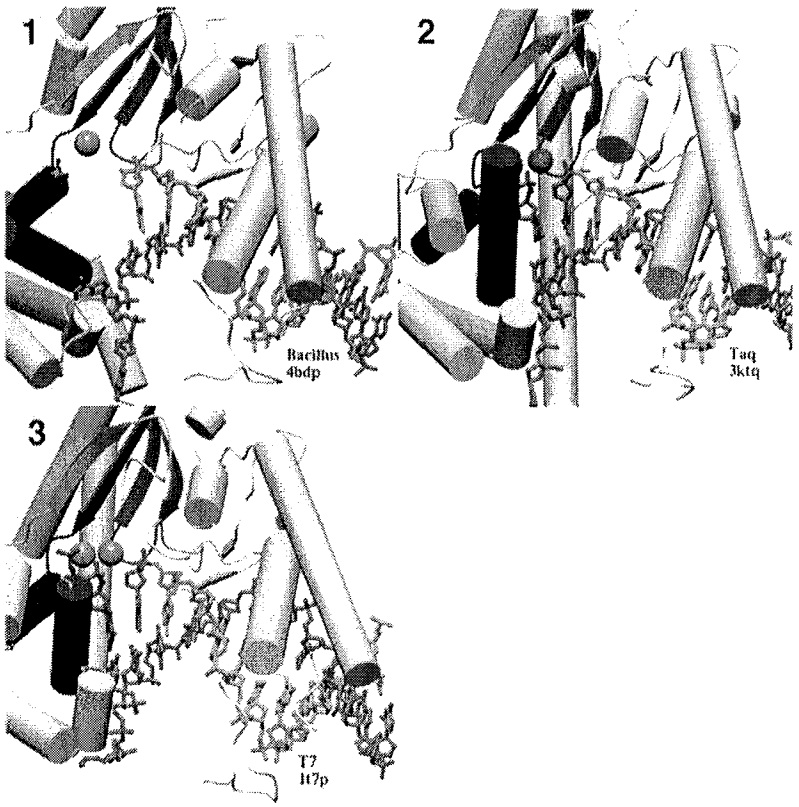


Figure 4. Active sites of *Bst* Pol I (Kiefer *et al.* (1998), panel 1), *Taq* Pol I (Li *et al.* (1998), panel 2), and T7 Pol (Doublie *et al.* (1998), panel 3). The location of motifs A (red), B (green) and C (blue), along with other secondary structural elements are nearly identical within these diverse polymerases. In addition, binding to DNA by each of these polymerases distorts the conformation of the nucleic acid. Notably, the first template bases are "flipped" out of the helix axis. *Bacillus* polymerase (*Bst* Pol I; 4bdp) depicts conformation of the protein following nucleotide incorporation; thus, the protein is bound to DNA but not an incoming nucleotide. In this state, the polymerase is in an "open conformation" with the fingers subdomain (motif B) located away from the catalytic triad residues located in the palm subdomain (motifs A and C). The *Taq* Pol I (3ktq) and T7 pol (1t7p) structures show polymerases bound to DNA and an incoming nucleotide. In this state, polymerases adopt a "closed" conformation, with fingers subdomain motif B elements located near the palm subdomain.

6). We propose all three of the following steps are central to an "induced-fit" model for nucleotide incorporation.

(1) Structural elements within the fingers domain rotate towards the 3′ primer terminus, resulting in a "closed" structure.

(2) The template base rotates back into the helix axis (by $\geqslant 90°$).

(3) The base portion of the incoming nucleotide forms a Watson-Crick base-pair with the template base and the triphosphate portion forms metal-mediated ionic interactions with amino acid residues of the active site.

Four high resolution DNA polymerase structures of the ternary enzyme:DNA:dNTP complex, two in an "open conformation" (*Bst* Pol I and *Taq* Pol I) and two in a "closed" conformation (T7 pol and *Taq* Pol I) provide valuable insights into the structural basis for nucleotide incorporation. Two unique conformations of *Taq* Pol I:DNA:dNTP complex were discovered by Waksman and colleagues, following crystallization experiments with varying amounts of the incoming nucleotide. In the presence of a low concentration of the incoming nucleotide, the polymerase is in an open conformation (Figures 5(a) and 6(a)), and the *Taq* Pol I structure (Li *et al.*, 1998) is identical to that of *Bst* Pol I (which depicts the conformation of polymerase following dNTP incorporation and poised for the next round of synthesis (Kiefer *et al.*, 1998)). In both of these structures, the first template base is flipped out $\geqslant 90°$ away from the helix axis, and a highly conserved motif B tyrosine residue (Tyr671 of *Taq* Pol I and Tyr714 of *Bst* Pol I; Figure 7) is positioned exactly where the template base would be expected, on top of the first base-pair of the double-stranded DNA. Thus, Tyr671 maintains stacking interactions expected of the template base. In this open conformation of the *Taq* Pol:DNA:dNTP complex, the triphosphate portions are coordinated by Asp610 and Asp785, and the base portion of the incoming dNTP lies in a solvent exposed pocket and stacks on top of the first base-pair of the dsDNA, adjacent to Tyr671. Watson-

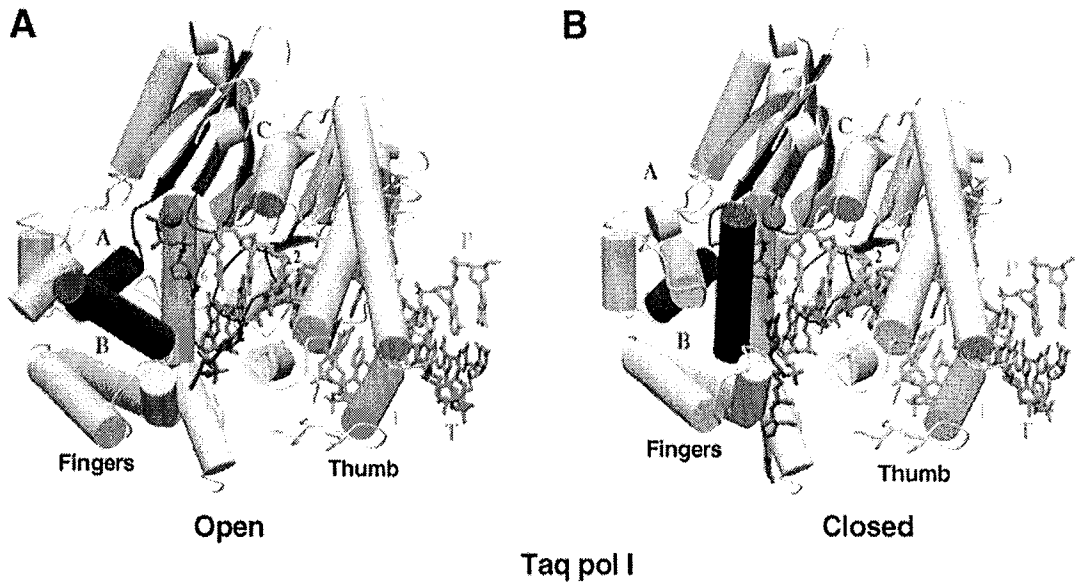


Figure 5. Location of conserved regions within *Taq* Pol I. Six conserved regions within DNA Pol I are shown in "open" (a) and "closed" (b) conformations of *Taq* Pol I. Region 1 (light green) is located within the thumb subdomain and interacts with the DNA minor groove. Region 2 (orange) is located within the palm subdomain and forms the "template grip" that interacts with the DNA near the minor groove. Motif A (red) interacts with an incoming divalent cation-dNTP complex. Motif B (dark green) also interacts with the incoming dNTP, but is located in the fingers subdomain and changes conformation to adopt a "closed" form during the nucleotide binding step. Motif C (blue), along with motif A, is located in the palm subdomain; both of these motifs harbor the catalytic triad carboxylate containing amino acids. Region 6 (green-yellow) is located in the palm subdomain and interacts with the template strand near the site of "base flipping". The primer strand is shown in yellow, the template strand is in cyan, the flipped out bases are colored in purple, and the incoming ddNTP is colored in magenta.

Crick base-pairing interactions are not observed initially, as the template base is flipped out.

At a higher concentration of the incoming nucleotide, the polymerase adopts a "closed" state (Li *et al.*, 1998) (Figures 5(b) and 6(b)). In this conformation: (a) the template base is "flipped" $\geqslant 90°$ into the helix axis; (b) there is a Watson-Crick base-pairing with the incoming nucleotide; (c) Phe667 of O-helix forms stacking interaction with the incoming dNTP; and (d) the fingers subdomain changes conformation to almost completely engulf the incoming nucleotide substrate. This conformational change affects significantly the position of motif B amino acid residues (O-helix) located within the fingers subdomain, and occurs concomitantly with a 40° rotation of the O-helix towards the catalytic site amino acid residues Asp610 and Asp785. Following these conformation changes, the base portion of the incoming nucleotide packs against two hydrophobic planar amino acid residues (Tyr671 and Phe667) located in motif B, and the ribose portion of the nucleotide packs against Ile614 and the aliphatic portion of Glu615. Thus, a hydrophobic pocket surrounds the base and ribose portions of the incoming dNTP. The negatively charged triphosphate portion of the dNTP, following the conformational change, interacts with two basic (positive charged) amino acid residues (Lys663 or Arg659) in motif B. Two acidic side-chains, Asp610 and Asp785, in motif A and C, respectively, also participate in metal-mediated interactions with the triphosphate group; these amino acid residues are directly involved in stabilizing the transition state during bond formation.

An induced-fit model for nucleotide incorporation should explain how the following three interactions with the incoming nucleotide are formed during dNTP binding: (1) hydrogen bonding with the template base; (2) stacking interactions with planar ringed amino acid residues; and (3) electrostatic interactions with negatively charged phosphate groups and charged side-chains. The *Taq* Pol I "open" and "closed" structures suggest that the triphosphate portion of the incoming dNTP initially binds near the catalytic residues Asp610 and Asp785 *via* metal coordination. We propose protein:dNTP-stacking interaction occurs as the O-helix Phe667 residue interacts with the base position of the incoming nucleotide. Development of this stacking interaction facilitates movement of the O-helix (motif B) towards motif A amino acid residues and displaces Tyr671, which in turn allows the template to rotate $\geqslant 90°$ back into the helix axis. If the incoming nucleotide is complementary to the base, then hydrogen bonding with the template base would further stabilize the dNTP. Stable stacking and base-pairing interactions would provide time for the tip of the finger (O-helix) to come in close proximity with the palm subdomain. This conformational change results in

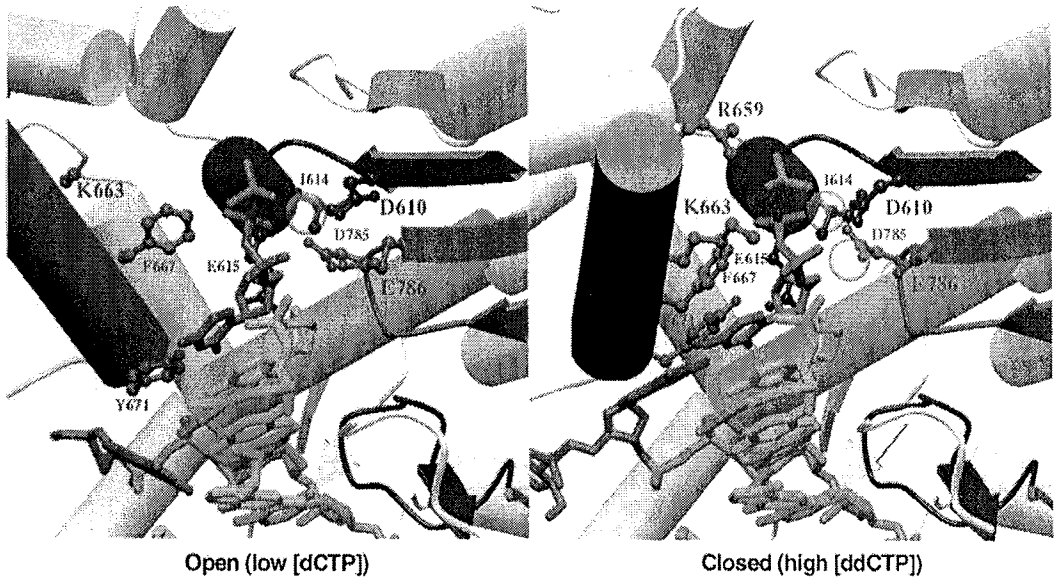


Figure 6. Conformational changes during dNTP binding. The Figure illustrates some of the major interactions and conformation changes during the dNTP-binding step. Initially, during dNTP binding (magenta), the template base (purple) is rotated out of the helix axis, thus cannot base-pair with the incoming nucleotide. The expected position of the template is taken up instead by a highly conserved Tyr (Y671) residue, which hydrogen bonds with the dNTP. In addition, motif B (O-helix; green) adopts an open conformation. Following the completion of the dNTP binding step: (1) conformation change by O-helix displaces Tyr671 from stacking with the base-pair; (2) the template base rotates into the helix axis; and (3) Watson-Crick base-pairing occurs. The conformation change by O-helix results in a generally hydrophobic pocket into which the base and ribose moiety bind and a hydrophilic pocket, which houses the triphosphate portion of the incoming dNTP. Following the conformation change the *para*-OH group of Tyr671 hydrogen bonds with the carboxylate of Glu615. In addition, this conformation change serves to bring the dNTP α-phosphate close to the primer 3′-OH group. The conformation of motif A (red), and C (blue) elements, as well as region 6 elements (light green) do not change significantly during the dNTP binding step.

hydrophobic interactions with the base and ribose portions of the incoming dNTP and hydrophilic interactions with the triphosphate group. Overall, the induced-fit model for nucleotide incorporation involving a conformation change of the O-helix and template rotation of $\geqslant 90°$ allows establishment of stacking interactions and serves to bring the dNTP α-phosphate close to the primer 3′-OH group, thus promoting chemical catalysis. This induced-fit mechanism for nucleotide selection would also restrict conformations and structures of the incoming nucleotide, promoting the efficiency of correct nucleotide incorporation. In summary, we propose that stable stacking interactions with the new base-pair, in conjunction with strong electrostatic interaction brought about by the movement of the tip of the O-helix into close proximity to the palm, results in a conformation that enhances base selection by DNA polymerases.

Role of the "base-flipping" mechanism

Crystal structures of several enzymes that process either normal or damaged bases (including glycosylases and methyl transferases) function by first rotating the designated base out of the DNA helix axis (Roberts, 1995). This base flipping mechanism involves a rotation around the phosphodiester bonds so that the base is flipped 180° out of helix axis and into the active site of the enzyme. All of the high-resolution DNA polymerase:DNA crystal structures indicate that the first template base is rotated out by $\geqslant 90°$ and the second template base is rotated by 180° (Figures 4-7). It is currently unknown whether this rotation in diverse DNA metabolizing enzymes involves an active mechanism that rotates the sugar phosphate backbone, or whether this rotation out of the helix axis occurs by passive diffusion and the rotated conformation is stabilized by specific active site amino acid residues. Within diverse Pol I structures complexed with DNA, amino acid residues within helix Q (Region 6 amino acids, specifically in *Taq* Pol Arg746 to Asp759) interact with the sugar-phosphate backbone near the first template base. The flipped out first template base is initially stabilized by stacking interactions between helices O and O1, and flipped out position of the second template base stacks with His676 of helix O1 within *Taq* Pol I (and Tyr719 of *Bst* Pol I).

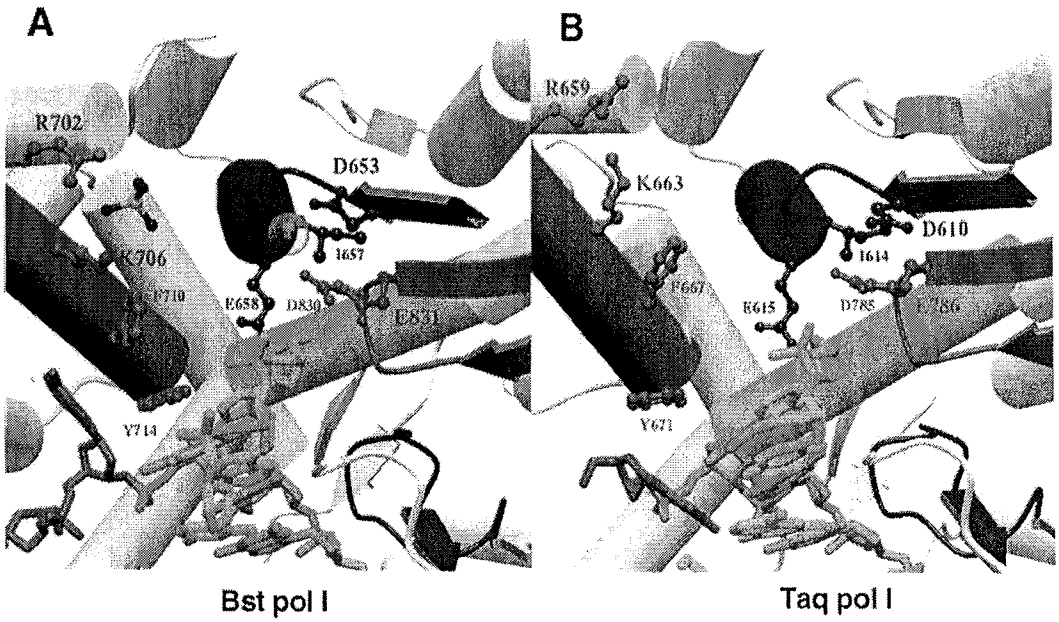


Figure 7. Conformation of the translocated complex. *Bst* Pol I complex (a) depicts the conformation of Pol I following nucleotide incorporation and translocation. The O-helix in this structure adopts an open conformation, Tyr714 (equivalent to Y671 of *Taq*) stacks with the newly synthesized base-pair, and the template base (purple) is rotated out of the helix axis. Faint electron density appears for the resulting pyrophosphate group near the basic O-helix residues. The overall conformation of *Bst* Pol I resembles that of *Taq* Pol I complexed with DNA in the absence of dNTP (b).

A mechanism for "base flipping" during DNA polymerization has several interesting implications. Such a mechanism could prevent −1 frame shift errors by hindering base-pairing to upstream templates, and thus enhance the fidelity of DNA replication. During the dNTP-binding step, the "flipping-in" mechanism allows establishment of stable stacking interactions between the new base-pair and planar side-chains within the polymerase active site (Tyr671 and Phe667). These stacking interactions help form a hydrophobic pocket that surrounds the base and ribose portions of the incoming dNTP and the template base. Thus, these base-stacking interactions contribute to the DNA polymerase base-pairing fidelity by allowing the enzyme to form a pocket that can recognize the shape of the incoming new base-pair, prior to the nucleotide incorporation step. Lastly, the torsion created by twisting the single-stranded template strand could be important for translocation (see below).

Nucleophilic attack

The catalytic site of diverse DNA polymerases is composed of three acidic amino acid residues with carboxylate side-chains (in *Taq* Pol I: Asp610, Asp785, and Glu786). The ternary closed *Taq* Pol I: DNA: dNTP structure shows that two of the carboxylate side-chains (Asp610 and Asp785) bind two metal ions (Li *et al.*, 1998); the precise role of the third carboxylate amino acid residues is unclear. Each metal (metal A and B; Figure 6(b)) is octahedrally coordinated and contains six ligands, four in one plane and one on each side of the plane. Metal A is coordinated by 3′ OH, the α-phosphate group, oxygen atoms of Asp610 and Asp785 side-chains, and two water molecules. Metal B is coordinated by: α-, β- and γ-phosphate groups, carboxylates of Asp610 and Asp785, and the carbonyl of Tyr611. The two metal sites are 3.8 Å apart, and each ligand is separated by 2.2 Å, a coordination symmetry consistent with octahedral coordination observed in many protein: dNTP structures. A similar two-metal coordination scheme also was observed in Pol β:DNA:dNTP (Pelletier *et al.*, 1994), T7 DNA polymerase: DNA:dNTP (Doublie *et al.*, 1998) and HIV RT:DNA:dNTP (Huang *et al.*, 1998) structures. The interaction of metal A with the α-phosphate is thought to enhance the electropositive character of the phosphorous atom; in addition, metal A interaction with primer 3′-OH is thought to facilitate deprotonation of the 3′-OH group (Brautigam & Steitz, 1998). Metal A coordination geometry brings the two reacting groups (the nucleophile 3′-OH and α-phosphate) in close proximity and facilitates the phosphoryl transfer reaction.

Pyrophosphate release and translocation

The phosphoryl transfer reaction leaves, as products, a primer elongated by a single nucleotide and a pyrophosphate group. Analysis of an "open" *E. coli* Pol I (Klenow) bound to PPi (Beese *et al.*, 1993b) and several "closed" Pol I structures

indicates that following nucleotide incorporation, the polymerases adopt an "open" conformation, translocate to next template position, and release PPi prior to beginning the next cycle of nucleotide incorporation. All Pol I class of enzymes not bound to both DNA and dNTP are in an "open" conformation. Within the "open" *E. coli* Pol I:PPi complex, PPi is complexed with the conserved Lys and Arg (within O-helix) motif B amino acid residues. Within the *Taq* Pol I:DNA:dNTP and T7 pol:DNA:dNTP "closed" complexes, the β and γ-phosphates of the dNTP interact with the O-helix conserved side-chains Arg659 and Lys663 and motif A amino acid residues Asp610 and Asp785, and the base and ribose portions of dNTP are bound in a hydrophobic pocket (Figure 6(b)). Thus, during the transition from a "closed" to "open" conformation, O-helix amino acid residues Arg659 and Lys663 can function to remove PPi away from the catalytic site. What causes the polymerase to adopt an "open" conformation? We propose conformational change into an "open" state is accompanied by a disruption of the stacking interactions following nucleotide incorporation and translocation. Consistent with this hypothesis, the structure of the *Bst* Pol I:DNA shows, following nucleotide incorporation and translocation, the enzyme is in an "open" conformation and there is no pocket that engulfs the incorporated nucleotide. In this structure, the PPi group has diffused away from the active site (Figure 7(a)).

Many of the Pol I:DNA interactions by region 1 and 2 amino acid residues involve the sugar-phosphate backbone along the minor groove. These non-specific interactions allow polymerase (or the DNA) to translocate freely in a spiral motion. Following nucleotide incorporation and translocation, the same interactions of the polymerase active site would need to be reformed at the new primer terminus. Thus, translocation needs to be halted when the primer 3′OH group is near the active site. Tyr714 within the translocated *Bst* Pol I:DNA complex (Tyr671 of *Taq* Pol I) projects into the DNA and stacks over the template base of the newly formed base-pair (Figure 7(a)). Following translocation, this interaction positions the primer: template at the catalytic site. The strain created by the DNA helix distortion (i.e. *A*-form DNA at the active site and 90-180° base flipping of the template strand) may help guide the direction of translocation (5′-3′) with respect to the primer. One can speculate that other DNA metabolizing enzymes, including helicases, may adopt a base flipping mechanism to simultaneously eliminate base-pairing and stacking interactions within the double helix during DNA unwinding. In addition, the distortion created by rotating the sugar phosphate backbone may guide the direction of DNA unwinding.

Function of Individual Amino Acid Residues in Polymerization

Structural data indicate that fewer than 10 out of over 50 amino acid residues in conserved motifs have direct interactions with the incoming dNTP substrate. For example, essential putative functions can be ascribed to amino acid residues Asp610 and Glu615 (of motif A), Tyr671, Phe667, Lys663, and Arg659 (of motif B), Asp785 and Glu786 (of motif C); however, the precise roles of all other amino acid residues within these conserved motifs are unclear. The function of different amino acid residues and side-chains in catalysis can be gleaned from a combination of kinetic measurements, site-specific mutagenesis, and random mutagenesis coupled with genetic selection. Site-specific mutagenesis has the advantage that one can assign a particular change in kinetics or substrate specificity to a distinct amino acid substitution. However, this approach is frequently limited in that catalytic efficiency of the enzyme is compromized, and the effects observed might reflect a global change in enzyme structure. Using random mutagenesis with positive genetic selection, one obtains mutant enzymes that exhibit catalytic efficiencies similar to the wild-type. However, many of the enzymes contain multiple substitutions that militate against the direct assignment of altered catalytic properties to specific substitutions. As a result, a combination of these approaches offers many advantages in structure-function studies. In considering the result so far obtained, we have chosen to analyze substitutions that occur in different motifs with the understanding that this simplification does not adequately consider distant interactions of amino acid substitutions.

Extensive number of site-directed mutagenesis studies, yielding enzymes with single amino acid substitutions have been done with the Klenow fragment, Pol β, and HIV RT. Kinetic analysis of these polymerases, which have a conserved amino acid residue substituted by a neutral residue, show these mutants frequently have low activity, and some mutants have a unique substrate specificity. For example, Joyce and colleagues (Minnick *et al.*, 1999) replaced 28 conserved side-chains individually within the *E. coli* Pol I (Klenow fragment) active site to alanine (or, in one example, to leucine), and tested each of the mutants for activity and fidelity. In these experiments, the majority of mutants had a 5- to 100-fold reduction in activity relative to wild-type using gapped DNA as a template. Only four of the 28 mutants tested exhibited low fidelity. Specifically, E710A within motif A (equivalent to E615A in *Taq* Pol I), Y766A (Y671A) within motif B, Arg668 (Arg573) and Arg682 (Arg587) within region 2, and N845A (Asn750) of region 6 have low fidelity. These results suggest fidelity is determined by very few conserved amino acid interactions. In addition, a complete deletion of region 1 amino acid residues (located at the tip of the thumb subdomain and binds within

the minor groove of the DNA duplex) results in polymerases with low activity, processivity, and high propensity for +1 frame-shift errors (Minnick *et al.*, 1996). Together, these site-directed mutagenesis and deletion mutagenesis studies have verified that many substitutions of conserved amino acid residues lower enzyme activity, and in some cases, alter substrate specificity.

Applied molecular evolution of the polymerase active site

To determine the contribution of specific active site residues to polymerase function and to evaluate the degree of plasticity within the polymerase active site *in vivo*, we randomly mutated 13 residues each within the two most highly conserved regions within DNA polymerases: motif A (amino acid residues 605-617 in *Taq* Pol I (Patel & Loeb, 2000a)) and motif B (amino acid residues 659-671 in *Taq* Pol I (Suzuki *et al.*, 1996b)). Briefly, these experiments involved replacement of nucleotide sequence encoding amino acid residues with a partially randomized sequence, such that each amino acid can be altered to potentially any of the other 19. When coupled with a stringent selection scheme, one can determine the nature of allowable amino acid substitutions *in vivo* after sequencing selected mutants. Functional mutants were selected by complementation of *E. coli recA718 polA12*, a strain that contains a temperature-sensitive mutation in the *polA* gene and can be propagated at 30 °C, but not at 37 °C (Sweasy & Loeb, 1992; Uyemura & Lehman, 1976). *Taq* Pol I can fully restore the temperature-sensitive phenotype such that *E. coli recA718 polA12* harboring *Taq* Pol I exhibits a 100% survival rate at 37 °C relative to 30 °C (Patel & Loeb, 2000a).

Following transformation of the motif A library into Pol I deficient *E. coli*, we obtained 8000 active mutants, of which 291 were characterized for activity, fidelity and ability to incorporate ribonucleotides (Patel & Loeb, 2000a). Sequence analysis of all 291 selected active *Taq* Pol I clones indicated that 8 out of 13 motif A amino acid residues tolerate a wide spectrum of substitutions, four amino acid residues tolerate conservative substitutions and only one amino acid residue (Asp610) is immutable. Comparison of a high-resolution *Taq* Pol I bound to DNA and dNTP complex shows that the immutable residue (Asp610) functions to coordinate the metal-mediated catalysis reaction, leading to the incorporation of the incoming nucleotide. Those residues that tolerate predominantly conservative substitutions stabilize tertiary structure. For example, Glu615, which hydrogen bonds with O-helix side-chain Tyr617, can only be substituted by an Asp, and Tyr611, which projects into a hydrophobic pocket, can only be replaced by other planar amino acid residues. Most other motif A amino acid residues can be substituted to a wide range of amino acids that differ in size, shape and charge, frequently without compromising activity. Interestingly, the conserved amino acid residues from our random mutagenesis experiments correlate well with motif A consensus sequence found in nature (Figure 3).

Transformation into *E. coli recA718 polA12* of motif B library containing random substitutions at amino acid residues 659-671 also yielded an extensive library of highly active DNA polymerases that contained diverse amino acid substitutions (Suzuki *et al.*, 1996b). Sequence analysis of the selected active clones showed only two of the 13 amino acid residues are immutable (Arg659 and Lys663), two tolerate conservative substitutions (Phe667 and Tyr671), and all others tolerate a wide range of substitutions. Structural analysis shows the immutable and conservatively substitutable amino acid residues interact with the incoming dNTP. These results are largely consistent with sequence alignment data, which show within prokaryotic and bacteriophage Pol I polymerases only motif B amino acid residues Arg659, Lys663, Gly668, and Tyr671 are conserved, and Phe667 tolerates only conservative substitutions. The random mutagenesis data suggest only 3 out of 26 active site (motif A and B) residues analyzed are required for function *in vivo*, six can tolerate only conservative substitutions, and all other residues are highly substitutable. This inherent plasticity of the *Taq* Pol I active site is consistent with Pol I consensus sequence found in nature (Figure 3). Taken together, these data indicate that the DNA polymerase active site is highly mutable.

Altered substrate specificity

DNA polymerases have an active site architecture that must specifically configure to and incorporate each of the four deoxynucleoside triphosphates, while taking direction from templates with diverse nucleotide sequences. In addition, the active site must exclude altered nucleotides produced during cellular metabolism. The amino acid residues of motifs A and B are positioned to have a potentially important contribution toward DNA polymerase fidelity and substrate specificity. The effect of amino acid substitutions within the active site of *Taq* DNA polymerase was analyzed using over 300 active mutants. Mutant libraries were tested for base-pair fidelity and for the incorporation of ribonucleotides.

Base-pairing fidelity

The conformational changes at the catalytic site of DNA polymerases that guarantee the high fidelity of nucleotide incorporation have not been adequately established. These changes must accommodate each of the complementary, yet structurally different base-pairings. One model that addresses the fidelity requirements is that during nucleotide incorporation, stable base stacking interactions and base-pairing allow time for a confor-

mational change, which brings the α-phosphate towards the primer 3′-OH (Patel *et al.*, 1995). If a non-complementary nucleotide were to bind, then the base-pairing and stacking interacting would be disrupted, and as a result, the conformational change would be retarded. In this model, low fidelity can be conferred by mutations that interfere with the transitions between "open" and "closed" conformations. Two types of mutations reduce fidelity: (1) mutations that stabilize the closed conformation, even in the absence of proper base-pairing and stacking; and (2) mutations that "widen" the hydrophobic pocket to accommodate binding of non-complementary nucleotides. Substitutions within the O-helix that confer low fidelity include Ala661Glu and Thr664Arg (Suzuki *et al.*, 1997, 2000). These residues are located near the distal portion of the fingers subdomain, adjacent to amino acid residues that interact with the triphosphate group (Arg659 and Lys663) and away from the base stacking residues (Tyr671 and Phe667) positioned near the palm subdomain. These amino acid substitutions confer low fidelity by stabilizing the closed conformation. Substitutions within motif A that confer low fidelity include hydrophilic substitutions at position 614 (Patel *et al.*, 2001). These mutants exhibit a tenfold higher efficiency of misinserting bases, as well as at least tenfold higher efficiency in extending mismatches. Following PCR, *Taq* Pol I mutants containing hydrophilic residue at position 614 exhibit up to 100-fold higher error rates relative to the wild-type enzyme and efficiently catalyze both transition and transversion errors. In addition to conferring low base-pairing fidelity, hydrophilic substitutions for Ile614 also allow mutant *Taq* Pol I to bypass blocking template lesions such as abasic site and vinyl chloride alkylation product ethenoA. Hydrophilic substitutions for Ile614 presumably confer a low fidelity by "widening" the dNTP-binding pocket to accommodate unusual template base and incoming nucleotide structures (for further discussion, see Patel *et al.*, 2001).

Incorporation of ribonucleotides

In cells, the concentration of ribonucleotides is orders of magnitude greater than that of the corresponding deoxynucleotides. As a result, DNA polymerases evolved mechanisms to prevent the incorporation of ribonucleotides in DNA (Joyce, 1997). We analyzed this property by testing 291 "motif A mutant enzymes" for the ability to synthesize RNA. Twenty-three different mutant polymerases containing substitutions in one of two amino acid residues were identified that incorporated ribonucleotides at a rate approaching 10^3-fold greater than that of wild-type *Taq* DNA polymerase (Patel & Loeb, 2000b). Many of the 23 ribonucleotide incorporating mutants contained multiple substitutions but could be divided into two major classes: (1) Those encoding a hydrophilic substitution at Ile614; and (2) those that encode a Glu615Asp substitution. Kinetic analysis of some of the mutants shows each incorporates ribonucleotides at an efficiency (k_{cat}/K_m) 1000 times higher relative to the wild-type enzyme. Analysis of the *Taq* Pol I structure model bound with DNA and rNTP shows the ribose ring of the rNTP interacts with Ile614 (specifically, with methyl group on the β-carbon). In addition, Ile614 amino acid residues in an important region of the protein located between an α-helix and a β-strand. Substitutions of Ile614, which is a highly mutable residue, may allow rNTP binding and incorporation by disrupting the shape of the hydrophobic pocket that interacts with the base and ribose portions of the incoming dNTP. The aliphatic portion of Glu615 also forms a portion of this hydrophobic pocket. Thus, Glu615Asp substitution may similarly disrupt the overall conformation of the pocket to allow ribonucleotides incorporation; however, this disruption is not significant enough to alter base-pairing fidelity of *Taq* Pol I. Taken together, these data indicate that 3 out of the 13 motif A amino acid residues (Ile614, Glu615, and Arg617), when substituted, altered biochemical properties of *Taq* Pol I, while many other substitutions at the other amino acid residues appear to be neutral in biochemical properties.

Evolution of the DNA Polymerase Active Site

Our genetic selection protocol allows isolation of mutant polymerases that retain a high DNA polymerase activity. Bacteria dependent on these over 300 mutant active polymerases can be grown under logarithmic conditions in liquid broth at 37 °C (prior to plasmid isolation and protein purification) or as colonies in solid agar at 37 °C (>50 generations) without significant variations in growth kinetics. Thus, bacteria dependent on mutant enzymes for survival are fit to replicate repetitively. In addition, many enzymes containing substitutions within motif A exhibit wild-type-like DNA-dependant DNA polymerase biochemical activity. The random mutagenesis data suggest only 3 out of 26 active site (motif A and B) residues analyzed are required for function *in vivo*, and another three or four amino acid residues, when substituted, alter the substrate specificity of the enzyme. This high level of substitutability of evolutionary conserved amino acid residues in the polymerase active site indicates multiple sequences can encode functioning polymerases.

This inherent plasticity of the DNA polymerase active site suggests DNA polymerases from geographically isolated populations of *E. coli* should contain diverse amino acid substitutions. Specifically, we expect after 10^8 years of evolution (Ochman & Wilson, 1988), dividing >100 times each year (Gibbons & Kapsimalis, 1967) at a mutation rate of 10^{-5} per nucleotide/division (in

mutators; Mao *et al.*, 1997; Oliver *et al.*, 2000; Taddei *et al.*, 1997) to 10^{-9} per nucleotide/division (in non-mutators) (Drake, 1991), *E. coli polA* genes might contain silent or neutral mutations at each codon position. However, analysis of *E. coli* strains isolated from 19 countries in four continents shows that each has a nearly identical amino acid and nucleotide sequence (>99% identity); nucleotides encoding motif A amino acid residues are 100% identical in all of these distinct *E. coli* strains. Consistent with the *E. coli* sequence data, two geographically distinct strains of *Rickettsia prowazekii* (strains B and Madrid E) have identical amino acid and nucleotide sequences for the entire *polA* gene (Andersson & Andersson, 1999). These data show that there is a high sequence identity in the *polA* gene of diverse *E. coli*. Comparison of *polA* sequences from individual species within the *Rickettsia* genus showed the greatest degree of nucleotide sequence identity (as judged by the presence of synonymous codons) occurs within the DNA polymerase and 5′ nuclease catalytic sites. For example, 24 amino acid residues within and flanking motif A are encoded by the identical codons in at least seven *Rickettsia* species. The frequency for this level of conservation for any one-codon position within the entire *polA* gene of *Rickettsia* is 0.8. Thus, the probability of 24 contiguous amino acid residues being encoded by the same codon by random chance is <0.005 ($=0.8^{24}$).

In summary, we find strains of identical species, isolated from geographically diverse areas have nearly the same nucleotide sequence within the *polA* gene, and species within the same genus have nearly identical active site nucleotide sequence, suggesting the *polA* gene evolves at a very slow rate. Thus, we are presented with two surprising and seemingly contradictory findings: DNA Pol I is highly plastic and allows multiple amino acid substitutions within its catalytic site, yet the *polA* gene is highly stable in nucleotide sequence within species. It is always conceivable that the wild-type amino acid sequence has a selective advantage over prolonged evolution; however, it seems improbable that selection would also prevail at the nucleotide sequence level. As an alternative, we propose that genetic transfer mechanisms maintain homogeneous sequences. Specifically, we propose the inherent plasticity of DNA polymerase enables tolerance of the high mutation burden during adverse conditions characterized by selection of mutators (Mao *et al.*, 1997) and facilitates the generation of beneficial mutations with a short-term selective advantage. Following successful survival through periods of adverse conditions, the wild-type sequence (one that is fit and the most prevalent) can be generated through genetic transfer, a process involving non-reciprocal recombination by substitution with wild-type sequences. Selection for a fit sequence could potentially be the driving force for genetic transfer. This model suggests that genetic transfer can function to maintain sequence homogeneity of DNA polymerase I and presumably other enzymes and could account for the dichotomy between inherent amino acid substitutability within diverse enzymes and the constancy of the nucleotide sequence found in nature. At the very least, the plasticity of the active site of DNA polymerase I, in contrast to the exceptional conservation of this region during evolution, suggests that there are mechanisms for maintaining nucleotide sequences during evolution that need to be explored.

Conclusions

Diverse prokaryotic species contain DNA polymerase I that are structurally identical and similar in sequence and in function. Furthermore, structural data indicate that very few (<10) amino acid residues within the highly conserved motifs A, B, and C have a direct role during nucleotide binding and incorporation. Random mutagenesis data indicate that only those residues that are important during catalysis and/or for protein folding need to be maintained; all other residues are mutable. While some substitutions alter the substrate specificity of the polymerase, many do not. Bacteria dependent on these mutated polymerases for survival are fit to replicate repetitively without significant variations in growth kinetics; however, the long-term fitness of bacteria harboring these mutant enzymes is unknown. Sequence analysis of the gene encoding Pol I shows strains of identical species isolated from geographically diverse areas have nearly the same nucleotide sequence within the *polA* gene, and species within the same genus have nearly identical active site nucleotide sequence, suggesting that the *polA* gene evolves at a very slow rate. These combined data provide the underpinning for the use of *E. coli* DNA polymerase I or *Taq* DNA polymerase I as a prototype for studying mechanisms and functions of DNA polymerases in diverse organisms.

References

Andersson, J. O. & Andersson, S. G. (1999). Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* **16**, 1178-1191.

Battista, J. R. (1997). Against all odds: the survival strategies of *Deinococcus radiodurans*. *Annu. Rev. Microbiol.* **51**, 203-224.

Beckman, R. A. & Loeb, L. A. (1993). Multi-stage proofreading in DNA replication. *Quart. Rev. Biophys.* **26**, 225-331.

Beese, L. S., Derbyshire, V. & Steitz, T. A. (1993a). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science*, **260**, 352-355.

Beese, L. S., Friedman, J. M. & Steitz, T. A. (1993b). Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry*, **32**, 14095-14101.

Brandis, J. W., Edwards, S. G. & Johnson, K. A. (1996). Slow rate of phosphodiester bond formation accounts for the strong bias that *Taq* DNA polymer-

ase shows against 2′,3′-dideoxynucleotide terminators. *Biochemistry*, **35**, 2189-2200.

Brautigam, C. A. & Steitz, T. A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54-63.

Bryant, F. R., Johnson, K. A. & Benkovic, S. J. (1983). Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry*, **22**, 3537-3546.

Chien, A., Edgar, D. B. & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127**, 1550-1557.

Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* **3**, 461-467.

Deluca, P. & Cairns, J. (1969). Isolation of *Escherichia coli* strains with a mutation affecting DNA polymerase. *Nature*, **224**, 1164-1166.

Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution (see comments). *Nature*, **391**, 251-258.

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA*, **88**, 7160-7164.

Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA Repair and Mutagenesis*, ASM Press, Washington DC.

Friedberg, E. C., Feaver, W. J. & Gerlach, V. L. (2000). The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance. *Proc. Natl Acad. Sci. USA*, **97**, 5681-5683.

Gibbons, R. J. & Kapsimalis, B. (1967). Estimeates of overall growth rate of intestinal microflora of hamsters, guinea pigs and mice. *J. Bacteriol.* **93**, 510-512.

Gutman, P. D., Fuchs, P., Ouyang, L. & Minton, K. W. (1993). Identification, sequencing, and targeted mutagenesis of a DNA polymerase gene required for the extreme radioresistance of *Deinococcus radiodurans*. *J. Bacteriol.* **175**, 3581-3590.

Gutman, P. D., Fuchs, P. & Minton, K. W. (1994). Restoration of the DNA damage resistance of *Deinococcus radiodurans* DNA polymerase mutants by *Escherichia coli* DNA polymerase I and Klenow fragment. *Mutat. Res.* **314**, 87-97.

Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drog resistance. *Science*, **282**, 1669-1675.

Joyce, C. M. (1997). Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc. Natl Acad. Sci. USA*, **94**, 1619-1622.

Joyce, C. M. & Grindley, N. D. (1984). Method for determining whether a gene of *Escherichia coli* is essential: application to the polA gene. *J. Bacteriol.* **158**, 636-643.

Joyce, C. M., Kelley, W. S. & Grindley, N. D. (1982). Nucleotide sequence of the *Escherichia coli* polA gene and primary structure of DNA polymerase I. *J. Biol. Chem.* **257**, 1958-1964.

Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal (see comments). *Nature*, **391**, 304-307.

Kim, Y., Eom, S. H., Wang, J., Lee, D.-S., Suh, S. W. & Steitz, T. A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*, **376**, 612-616.

Kornberg, A. & Baker, T. (1992). *DNA Replication*, 2nd edit., W.H. Freeman and Co., New York.

Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallog.* **24**, 946-950.

Kuchta, R. D., Benkovic, P. & Benkovic, S. J. (1988). Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry*, **27**, 6716-6725.

Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.

Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V. & Cech, T. R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase (see comments). *Science*, **276**, 561-567.

Mao, E. F., Lane, L., Lee, J. & Miller, J. H. (1997). Proliferation of mutators in a cell population. *J. Bacteriol.* **179**, 417-422.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0, a program for photorealistic molecular Graphics. *Acta Crysallog. sect. D*, **50**, 869-873.

Minnick, D. T., Astatke, M., Joyce, C. M. & Kunkel, T. A. (1996). A thumb subdomain mutant of the large fragment of *Escherichia coli* DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. *J. Biol. Chem.* **271**, 24954-24961.

Minnick, D. T., Bebenek, K., Osheroff, W. P., Turner, R. M., Jr, Astatke, M. & Liu, L. *et al.* (1999). Side-chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067-3075.

Ochman, H. & Wilson, A. C. (1988). Evolution in bacteria: evidence for universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**, 74-86.

Oliver, A., Canton, R., Campo, P., Baquero, F. & Blazquez, J. (2000). High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*, **288**, 1251-1254.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762-766.

Patel, P. H. & Loeb, L. A. (2000a). DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl Acad. Sci. USA*, **97**, 5095-5100.

Patel, P. H. & Loeb, L. A. (2000b). Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* **275**, 40266-40272.

Patel, S. S., Wong, I. & Johnson, K. A. (1991). Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry*, **30**, 511-525.

Patel, P. H., Jacobo-Molina, A., Ding, J., Tantillo, C., Clark, A. D., Jr & Raag, R., *et al.* (1995). Insights into DNA polymerization from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry*, **34**, 5351-5363.

Patel, P. H., Kawate, H., Adman, E., Ashbach, M. & Loeb, L. A. (2001). A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *J. Biol. Chem.* **276**, 5044-5051.

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP (see comments). *Science*, **264**, 1891-1903.

Roberts, R. J. (1995). On base flipping. *Cell*, **82**, 9-12.

Steitz, T. A. (1999). DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395-17398.

Suzuki, M., Baskin, D., Hood, L. E. & Loeb, L. A. (1996a). Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites *in vivo* with the crystal structure. *Proc. Natl Acad. Sci. USA*, **93**, 9670-9675.

Suzuki, M., Christians, F. C., Kim, B., Skandalis, A., Black, M. E. & Loeb, L. A. (1996b). Tolerance of different proteins for amino acid diversity. *Mol. Div.* **212**, 111-118.

Suzuki, M., Avicola, A. K., Hood, L. & Loeb, L. A. (1997). Low fidelity mutants in the O-helix of *Thermus aquaticus* DNA polymerase I. *J. Biol. Chem.* **272**, 11228-11235.

Suzuki, M., Yoshida, S., Adman, E. T., Blank, A. & Loeb, L. A. (2000). *Thermus aquaticus* DNA polymerase I mutants with altered fidelity: interacting mutations in the O helix. *J. Biol. Chem.* **275**, 32728-32735.

Sweasy, J. B. & Loeb, L. A. (1992). Mammalian DNA polymerase β can substitute for DNA polymerase I during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **267**, 1407-1410.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H. & Godelle, B. (1997). Role of mutator alleles in adaptive evolution. *Nature*, **387**, 700-702.

Uyemura, D. & Lehman, I. R. (1976). Biochemical characterization of mutant forms of DNA polymerase I from *Escherichia coli*. I. The polA12 mutation. *J. Biol. Chem.* **251**, 4078-4084.

Edited by J. H. Miller

(Received 1 November 2000; received in revised form 1 February 2001; accepted 14 March 2001)